# Colloquium

# Bacteriophage T4 gene 41 helicase and gene 59 helicase-loading protein: A versatile couple with roles in replication and recombination

Charles E. Jones*, Timothy C. Mueser†, Kathleen C. Dudas‡, Kenneth N. Kreuzer‡, and Nancy G. Nossal*§

*Laboratory of Molecular and Cellular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0830; †Department of Chemistry, University of Toledo, 2801 West Bancroft Street, Toledo, OH 43606; and ‡Department of Microbiology, Duke University Medical Center, Durham, NC 27710

Bacteriophage T4 uses two modes of replication initiation: origin-dependent replication early in infection and recombination-dependent replication at later times. The same relatively simple complex of T4 replication proteins is responsible for both modes of DNA synthesis. Thus the mechanism for loading the T4 41 helicase must be versatile enough to allow it to be loaded on R loops created by transcription at several origins, on D loops created by recombination, and on stalled replication forks. T4 59 helicase-loading protein is a small, basic, almost completely α-helical protein whose N-terminal domain has structural similarity to high mobility group family proteins. In this paper we review recent evidence that 59 protein recognizes specific structures rather than specific sequences. It binds and loads the helicase on replication forks and on three- and four-stranded (Holliday junction) recombination structures, without sequence specificity. We summarize our experiments showing that purified T4 enzymes catalyze complete unidirectional replication of a plasmid containing the T4 *ori(uvsY)* origin, with a preformed R loop at the position of the R loop identified at this origin *in vivo*. This replication depends on the 41 helicase and is strongly stimulated by 59 protein. Moreover, the helicase-loading protein helps to coordinate leading and lagging strand synthesis by blocking replication on the *ori(uvsY)* R loop plasmid until the helicase is loaded. The T4 enzymes also can replicate plasmids with R loops that do not have a T4 origin sequence, but only if the R loops are within an easily unwound DNA sequence.

**T**he bacteriophage T4-infected cell is a replication factory designed for the rapid production of multiple copies of its genome. T4 replication does not need to be coordinated with cell division or a cell cycle. Instead, the T4 replication strategy is optimized for quickly producing and packaging its DNA. The phage uses two modes of replication initiation: origin-dependent replication early in infection and recombination-dependent replication at later times. Because the 168-kb linear phage genome is terminally redundant and circularly permuted, the end of one DNA molecule is homologous to the middle of another. Thus the products of early origin-dependent replication can invade each other to form D loops for recombination-initiated replication. The same complex of relatively few T4-encoded replication proteins is responsible for DNA synthesis in both modes of replication (reviewed in refs. 1–4).

The genes for each of the T4 replication proteins have been cloned, and the functions of the proteins have been characterized by *in vitro* reactions on model templates (Fig. 1) (reviewed in ref. 5). T4 DNA polymerase (gene 43), which catalyzes DNA synthesis on both leading and lagging strands, is attached to a clamp protein (gene 45) that is loaded by the complex of the gene 44 and 62 proteins. In the presence of the T4 gene 32 single-stranded DNA binding protein, T4 DNA polymerase, the clamp, and the clamp loader are sufficient for slow strand displacement synthesis of the leading strand. The 5′ to 3′ gene 41 helicase unwinds DNA ahead of the fork and increases the elongation rate more than 10-fold to 400 nt/sec, comparable to that *in vivo*. Although the helicase can load on nicked and forked DNA by itself, its loading is greatly accelerated by the 59 helicase-loading protein. Lagging strand fragments are initiated by pentamer RNA primers, whose synthesis requires both the helicase and gene 61 primase. These primers ultimately are removed by a T4 encoded 5′ to 3′ nuclease (T4 RNase H), and after gap repair, the adjacent fragments are joined by T4 DNA ligase. T4 type II topoisomerase (genes 39, 52, and 60) is required for replication *in vitro* when the parental duplex is a covalently closed circle.

In tightly controlled replication systems like *Escherichia coli* oriC or yeast ARS I, sequence-specific origin binding proteins help to load the helicase at the origin (6). In contrast, the seven known T4 replication origins, identified by a variety of methods, do not share DNA sequences (1, 3). The four best-characterized origins, *oriA*, *oriE*, *ori(uvsY)(F)*, and *ori(34)(G)*, each have a transcription promoter, whose transcript is thought to act as the primer for leading strand synthesis. With the exception of the repEB and repEA proteins at *oriE* (7), there is no evidence for T4-encoded proteins that are required for initiation at a specific origin. Consequently, the mechanism for loading the T4 41 helicase must be versatile enough to allow it to load on R loops created by transcription at several origins, on D loops created by recombination, and on stalled replication forks.

In this paper we review recent evidence demonstrating that the T4 gene 59 helicase-loading protein recognizes specific structures rather than specific sequences. It binds and loads the helicase on replication forks and on three- and four-stranded (Holliday junction) recombination structures, without sequence specificity (8, 9). We summarize our experiments showing that the purified T4 enzymes catalyze complete unidirectional replication of a plasmid containing T4 *ori(uvsY)*, with a preformed R loop at the position of the R loop identified *in vivo* (10). This replication depends on the

helicase and is strongly stimulated by 59 protein. Moreover, the helicase-loading protein helps to coordinate leading and lagging strand synthesis by blocking replication on the *ori(uvsY)* R loop plasmid until the helicase is loaded.

## T4 Gene 59 Helicase-Loading Protein

T4 mutants in gene 59 are UV-sensitive and defective in repair, recombination, and the recombination-initiated replication predominant in the late stage of infection (11, 12). The initial characterization of purified 59 protein showed that it was a small, monomeric, and basic protein that was capable of binding both single- and double-stranded DNA. 59 Protein also was shown to interact specifically with 41 helicase and gene 32 single-stranded DNA binding protein in the presence and absence of DNA (13–16). Barry and Alberts (16) showed that 59 protein increased the rate of loading of the 41 helicase on both single-stranded and nicked duplex templates. Its ability to load the helicase on single-stranded DNA covered with 32 protein was particularly striking because 32 protein strongly inhibits DNA unwinding by the helicase without 59 protein.

**Structure of 59 Protein.** The crystal structure of the full-length (217 residues) 59 helicase-loading protein [Protein Data Bank (PDB) code 1C1K; ref. 8] revealed a novel, almost entirely α-helical protein, with no similarity to the structures of other single-stranded DNA binding proteins (Fig. 2). Its 13 α-helices are divided into N and C domains of similar size. The single short β-sheet connects N-terminal residues 2–4 with residues 197–199 near the C terminus. There is a narrow groove between the two domains on the "top" of the protein. The surface of the protein is notable for the high density of basic and hydrophobic residues, which may be important for its DNA and protein interactions.

The sequence of 59 protein has only limited similarity to the sequences of functionally related proteins. However, the N domain of 59 protein has strong structural similarity to several members of the high mobility group (HMG) family proteins, including rat HMG1A and the LEF-1 lymphoid enhancer-binding factor (8). The HMG domain has been found in structure-specific nonhistone chromatin-binding proteins and sequence-specific transcription factors. Proteins with the HMG domain have been shown to bind to the minor groove of DNA, bending and partially unwinding the duplex. HMG1 and other family members bind and unstack cruciform DNA (reviewed in refs. 17–19). The structural alignment on the bottom of Fig. 2 shows the close superposition of 59 protein helices H1, H2, and H3 (residues 11–67) (blue) with residues 9–65 of the rat HMG1 structure (green) (PDB code 1AAB; ref. 20). The superposition of the duplex DNA from the LEF1-DNA complex
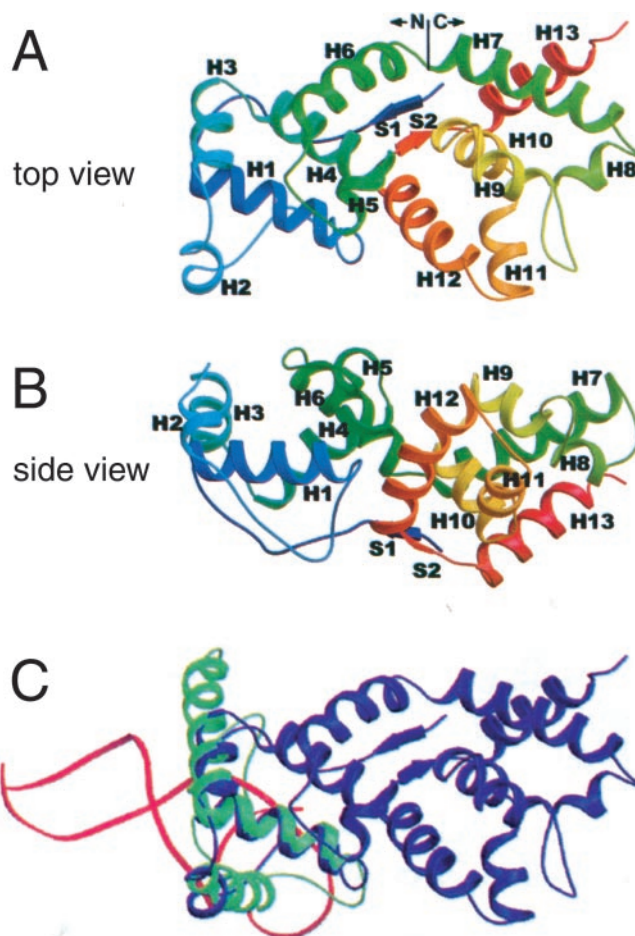


**Fig. 2.** Ribbon diagrams of the crystal structure of the T4 gene 59 helicase-loading protein showing its structural similarity with HMG proteins. (*A*) and (*B*) are, respectively, "top" and "side" view rainbow-colored ribbon diagrams of 59 protein (PDB 1C1K, ref. 8) from the blue N terminus to the red C terminus. S1–2 are β-sheets and H1–13 are α-helices. The line between H6 and H7 indicates the junction between the N and C domains of the protein. (*C*) Superposition of the H1, H2, and H3 α-helices of T4 59 helicase-loading protein (blue) and rat HMG1 protein (green, PDB 1AAB, ref. 20) and with the DNA (red) from the model of lymphoid enhancer-binding factor LEF1-DNA, (PDB 2LEF, ref. 21). Adapted from Mueser et al. (8).

(red) (PDB code 2LEF; ref. 21) suggests a possible orientation for duplex DNA on 59 protein.

**59 Protein Binds Preferentially to Forked DNA.** Although 59 protein binds to both single- and double-stranded DNA, it shows a clear preference for binding to DNA structures related to replication forks and recombination intermediates (8, 9). Our binding data, summarized in Fig. 3B, indicate that 59 protein binds strongly to fork DNA with either single- or double-stranded arms. These arms needed to be at least 6 nt long, but binding affinity was not increased by lengthening the arms from 12 to 18 nt. Whereas 59 protein bound to forks with completely duplex arms, it could not load the helicase unless there was a single-stranded gap of 10 nt on the fork arm corresponding to the lagging strand template. There was no requirement for a gap on the leading strand template. Three-way junction branched DNA, which is less flexible than fork and flap structures (22), was not bound by 59 protein. These observations are consistent with the role of the 41 helicase during DNA replication. The helicase moves 5′ to 3′ on the lagging strand template, opening the duplex ahead of the leading strand polymerase, and interacting with the primase to allow it to make the pentanucleotide primers that initiate lagging strand synthesis.
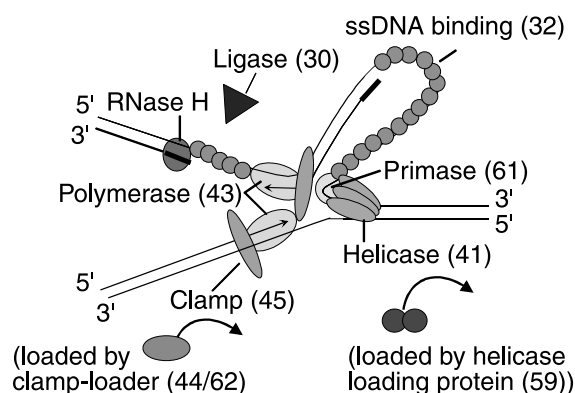


**Fig. 1.** Model of a replication fork with bacteriophage T4 proteins. Numbers indicate the name of the gene encoding each protein. See text for description of the functions of the proteins.
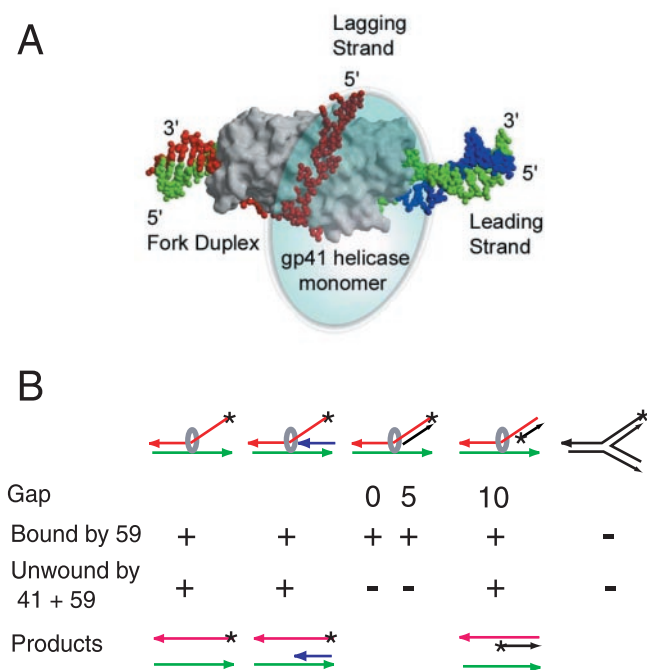
**Fig. 3.** Fork DNA structures bound by T4 59 helicase-loading protein and unwound by the 41 helicase loaded by 59 protein. (*A*) A speculative model of T4 gene 59 helicase-loading protein bound on a DNA replication fork. This model is taken from ref. 9 and was adapted from Mueser *et al.* (8). It is based on the distribution of hydrophobic and basic residues on the surface of 59 protein and on the assumption that the HMG-like region of its N domain binds and unstacks the duplex (red and green) ahead of the fork, as shown for HMG proteins (17, 19). The leading strand duplex [template (green)/primer (blue)] is docked on the bottom surface of the C domain. A long segment of single-stranded DNA, representing the lagging strand (red), traverses the shallow groove between the N and C domains. A helicase monomer (light blue oval) is proposed to bind to 59 protein between the lagging and leading strand arms (8). The other subunits of the hexameric helicase (not shown) would surround the lagging strand. (*B*) 59 protein binds to forks with single or duplex arms, but cannot load the helicase without a single-stranded gap of more than 5 nt on the lagging strand template. The fork DNA structures are drawn in the same orientation as the fork DNA on the model of 59 protein in *A*. The hexameric helicase (circle) is placed on the strand that would occupy the position of the fork lagging strand, if each of these DNAs binds to 59 protein as predicted by the model in *A*. Arrowheads represent the 3′ end of the substrates tested; * marks the position of $^{32}$P label in the substrates tested. [Reproduced with permission from ref. 9 (Copyright 2000, American Society for Biochemistry and Molecular Biology).]

Because the helicase-loading protein has much greater affinity for fork DNA with single- or double-stranded arms than for either single-stranded or simple duplex DNA, this small protein must have sites for the duplex region of the fork, as well as two arm sites that can accommodate either single- or double-stranded arms. Unfortunately, we do not yet have a structure for 59 protein bound to DNA. Fig. 3*A* shows a speculative model for 59 protein on a replication fork DNA that is based on the assumption that the "HMG-like" N domain binds the duplex ahead of the fork and holds the beginning of the fork arms in an open configuration (8). The leading strand duplex then could bind to a hydrophobic region on the bottom surface of the C domain, whereas the lagging strand template occupies the groove between the two domains on the top surface of the protein. If 59 protein binds as proposed in this model, it would separate the two fork strands and provide a docking area for 41 protein, which has been shown to surround the lagging strand (23). It also would be in position to bind to 32 protein coating the template for the lagging strand.
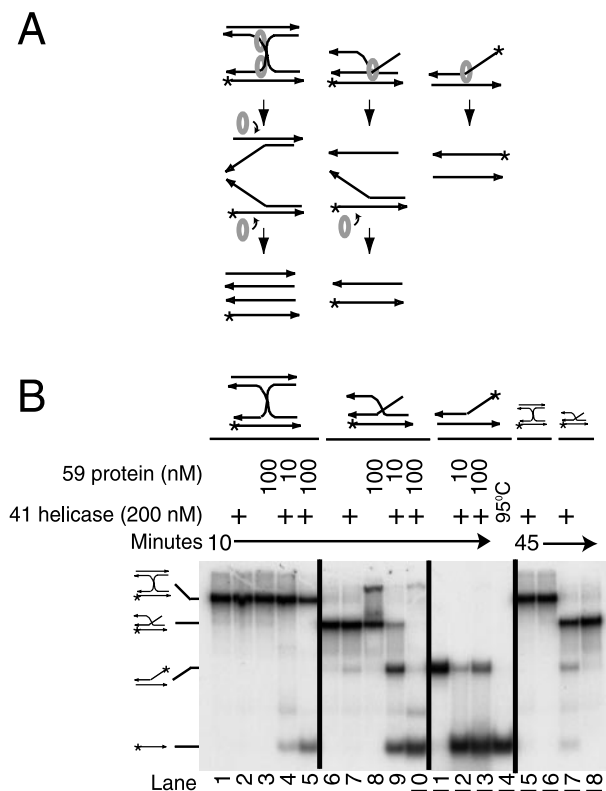


**Fig. 4.** T4 59 protein stimulates unwinding of cruciform DNA substrates by 41 DNA helicase. (*A*) Diagram showing the expected unwinding products and positions of the helicase (circle) on a four-stranded cruciform and three-stranded invasion structure, assuming that 59 protein binds to the fork portions of these structures and loads the helicase as predicted by the speculative model in Fig. 3*A*. Arrowheads represent the 3′ end of each strand; * marks the position of $^{32}$P label in the substrates tested. Strand exchange was not possible because the arms of the structures were not homologous. (*B*) Unwinding of cruciform and strand invasion structures. Helicase was present at a final concentration of 200 nM (monomer) where indicated. DNA was present at 10 nM. Adapted from ref. 9.

**Binding and Unwinding on Forks within Recombination Structures.** Both the 59 helicase-loading protein and the 41 helicase play vital roles in recombination. Gene 59 mutants and some gene 41 mutants have a recombination-defective phenotype (12, 24–26). Salinas and Kodadek (27) showed that after homologous pairing *in vitro* by the T4 uvsX, uvsY, and 32 proteins, both the 41 helicase and 59 helicase-loading protein were needed for extensive polar branch migration. Branch migration from circular single-stranded DNA annealed on a single-stranded extension of a complementary duplex is carried out by the combination of 41 helicase and the 59 and 32 proteins (28).

As anticipated because of the role of 41 helicase and 59 loading protein in recombination, 59 protein bound a three-stranded DNA strand invasion structure and four-stranded Holliday junction DNA (9) (Fig. 4). However, as indicated above (Fig. 3*B*), it did not bind three-way junction branch DNA. The Holliday junction and strand invasion structures were unwound by the helicase if 59 protein was present. In Fig. 4*A*, the strand invasion and cruciform structures are drawn in the same orientation as the fork DNA on the model of 59 protein in Fig. 3*A*. The hexameric helicase (circle) is shown on the strand that would occupy the position of the lagging strand, if these DNAs bind to 59 protein as predicted by the fork model.

In the three-stranded structure, the 3′ end of the invading strand has annealed to a complementary region of the duplex. 41 Helicase would be expected to bind to the 5′ end of this invading strand, because it is the only 5′-ended single strand in the

structure (Fig. 4*A*). The duplex that includes the 3′ end of this strand would be predicted by the model to be bound by the HMG-like N-terminal domain of 59 protein. In the experiment shown, the P³² label (indicated by * on Fig. 4*A*) was not on this invading strand. Thus helicase moving 5′ to 3′ on the invading strand would give unlabeled single-strand and labeled fork as the initial products. The labeled fork would be rapidly unwound to single strands by the helicase. In agreement with this prediction, both labeled fork and single strand were observed as products in the reaction with the lower concentration of 59 protein, whereas at the higher 59 protein concentration there was only labeled single-stranded product (Fig. 4*B*). This is the helicase reaction required for conservative recombination-initiated DNA replication and bubble-branch migration (see Fig. 7).

The observation that helicase can unwind the cruciform, which lacks single-stranded arms, suggests that there are enough accessible unpaired nucleotides at the cruciform junction to permit slow loading of the helicase (9). This would be expected if the HMG-like domain of 59 protein, like other HMG proteins, binds the cruciform in an unstacked open configuration.

The arms of the cruciform substrate in Fig. 4 do not share common sequences, so strand exchange was not possible in this experiment. If two helicases were loaded on the cruciform as shown on the diagram in Fig. 4*A*, the initial products would be two forks. Additional 41 helicase would be quickly loaded on these forks by 59 protein. Thus, the finding of single-stranded DNA as the only product (Fig. 4*B*) is not surprising, because the rate of unwinding of forks to single strands is so much faster than unwinding of the cruciform. If only one of the two helicases shown on the cruciform in Fig. 4*A* was present, no unwound product would have been detected. The single helicase would move 5′ to 3′ on the strand it encircles, opening the duplex ahead. Because the other ends of the strands forming this duplex are still bound in the four-stranded structure, the unwound duplex would reanneal rapidly. However, it remains to be determined whether unwinding of one arm of a cruciform by a single gene 41 helicase would be enough to facilitate strand exchange between homologous arms of a cruciform.

## Role of the Helicase and Helicase-Loading Protein During Origin-Dependent Replication

Gene 41 helicase is needed for both early origin-dependent T4 phage replication and replication of plasmids with the cloned T4 origins *ori(uvsY)* and *ori(34)*. The role of gene 59 helicase-loading protein in origin-dependent replication is less clear because amber mutations in gene *59* did not reduce early T4 phage DNA synthesis (29), or replication of *ori(uvsY)* plasmids (30), but did abolish plasmid replication via the recombination-dependent mode. Thus either 59 protein is necessary to load the helicase onto recombination intermediates, but not onto replication origins, or the gene *59* amber mutations are leaky, and a small amount of read-through 59 protein is sufficient for origin- but not recombination-initiated replication.

Two T4 origins, *ori(uvsY)* and *ori(34)*, were found to have a minimal sequence of about 100 bp that is sufficient for maximal levels of replication of plasmids containing these origins in T4-infected cells (31). Each of these minimal origins consists of a phage middle-mode promoter just upstream of a DNA-unwinding element (DUE) (32). T4 middle-mode promoters are transcribed by *E. coli* RNA polymerase that has been modified by addition of the T4 AsiA protein and must be activated by T4 MotA protein (33). Analysis of replicative intermediates during phage infection *in vivo* indicated that the RNA within a stable R loop at *ori(uvsY)* is used to prime leading-strand replication in one direction from the origin, and that assembly of a functional replication complex for the second direction is delayed significantly (34).

As an initial step toward reconstructing origin-initiated replication, we recently have shown that bacteriophage T4 proteins catalyze complete unidirectional replication of a plasmid,
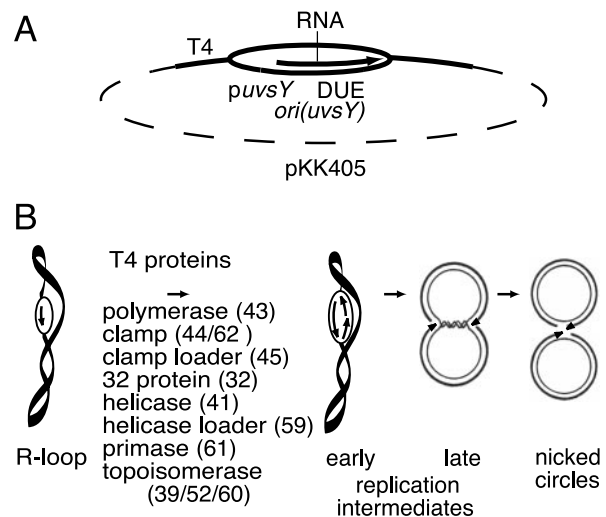


**Fig. 5.** Replication of a plasmid with a preformed R loop at the T4 *ori(uvsY)* replication origin by T4 replication proteins *in vitro*. (*A*) Diagram of the pKK405 plasmid (57), which contains a 1.35-kb fragment of T4 DNA (solid line) with the *ori(uvsY)* replication origin in a pBR322-based vector (dashed line). As indicated in the text, this origin contains the PuvsY transcription promoter adjoining a DUE. The R loop in the plasmid was formed by annealing a 104-b RNA corresponding to −5 to +99 relative to the transcription start site (10). (*B*) T4 replication proteins required for unidirectional replication from the pKK405 R loop plasmid, giving nicked circular products. [*B*, Reproduced with permission from ref. 10 (Copyright 2001, Elsevier Science).]

pKK405, with a preformed R loop at the T4 *ori(uvsY)* replication origin *in vitro* (10) (Fig. 5*A*). The 104-b RNA in the R loop serves as the primer for leading strand synthesis. The origin R loop plasmid was converted to two nicked circles in a reaction with T4 DNA polymerase, clamp, clamp loader, 32 protein, 41 helicase, 61 primase, 59 helicase-loading protein, and topoisomerase (Fig. 5*B*, and reactions 1 and 9 on the neutral gel, Fig. 6*B*). There were also some slower moving products, labeled late replication intermediates, which moved in the position expected for almost completed theta structures (35), as well as products just entering the gel, which are likely to be the result of rolling circle synthesis from the nicked circles. We showed that the nicked circular products were sealed if T4 RNaseH and DNA ligase are added to the reaction (10). As expected, leading strand synthesis stalled prematurely in the absence of T4 DNA topoisomerase.

Helicase was required for the synthesis of the nicked circular products from the R loop plasmid (10). When helicase activity was eliminated by omitting both the helicase and the loading protein, the products migrated as partially completed theta structures (early replication intermediates) (reaction 5, Fig. 6*B*). This limited synthesis depended on polymerase, clamp, clamp loader, and 32 protein, in agreement with the protein requirements for strand displacement synthesis (10). When the helicase-loading protein was omitted but the helicase was present (reaction 2), the DNA products were similar to those in the complete reaction, but the extent of replication was reduced about 4-fold. Thus helicase by itself can load on the replication bubble formed at the R loop, but loading is much more efficient with the loading protein. This is consistent with previous findings that 59 protein greatly stimulates, but is not essential for, strand displacement synthesis on nicked and forked duplex DNA (16, 36), and that there is some origin-initiated replication with T4 gene 59 mutants (1, 29).

When 61 primase was omitted (reaction 4 in Fig. 6*B*), there was a large increase in the synthesis of products with long single strands (10). The longest of these are likely formed by rolling circle synthesis from nicked circles. There also may be bubble-branch migration synthesis in which the leading strand polymerase moves continuously around the circle without enlarging the bubble because the
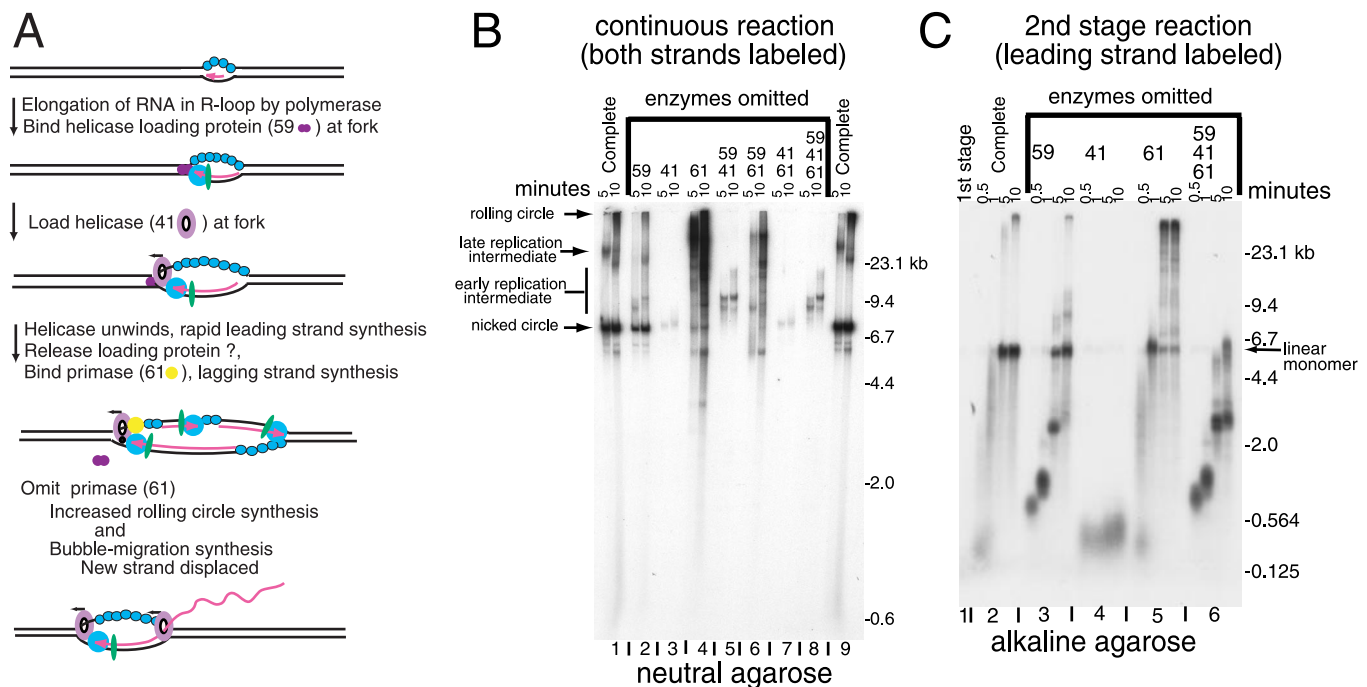
**Fig. 6.** Replication of the R loop plasmid with the T4 *ori(uvsY)* origin requires 41 helicase and is strongly stimulated by the 59 helicase-loading protein. (*A*) Model showing the roles of the 59 helicase-loading protein, 41 helicase, and 61 primase in unidirectional replication from T4 *ori(uvsY)*. 59 Protein binds ahead of the polymerase holoenzyme elongating the RNA in the R loop and loads the helicase essential for rapid leading strand synthesis. 59 Protein inhibits leading strand synthesis when there is no helicase. (*B*) Neutral agarose gel of the products of unidirectional continuous replication of the pKK405 *ori(uvsY)* R loop plasmid. The complete reactions contained the T4 replication proteins listed in Fig. 5*B*. 59 Helicase-loading protein, 41 helicase, and 61 primase were omitted as indicated. (*C*) Alkaline agarose gel of the replication of the pKK405 R loop plasmid in a two-stage reaction in which only the leading strand is labeled. The RNA was 3′ end-labeled by limited synthesis with dATP, dTTP, $^{32}$P-dCTP in a first-stage reaction that required T4 polymerase, clamp, clamp loader, and 32 protein. 41 Helicase, 59 helicase loader, 61 primase, and topoisomerase then were added along with dGTP and enough unlabeled dCTP to prevent further $^{32}$P-dCTP incorporation. The indicated enzymes were omitted in the second stage. T4 topoisomerase was present in all reactions. See text for description of the inhibition by 59 helicase-loading protein when 41 helicase is omitted and of the aberrant products made without the 61 primase. [*B* and *C*, Reproduced with permission from ref. 10 (Copyright 2001, Elsevier Science).]

single-stranded side of the bubble branch migrates to displace the newly synthesized leading strand. Bubble branch migration was first shown in replication reactions with the T4 polymerase holoenzyme, UvsX recombination protein, and Dda helicase (37). It also would be expected in these reactions with the 41 helicase and 59 protein, but without primase, because 41 helicase can unwind the invading strand from a three-stranded invasion structure (Fig. 4*B*) (9), and the combination of 59 helicase-loading protein, 41 helicase, and 32 single-stranded DNA binding protein catalyzes strand exchange (27, 28).

**59 Protein Helps to Coordinate Leading and Lagging Strand Synthesis by Blocking Replication in the Absence of the Helicase.** The 59 helicase-loading protein strongly inhibited replication when it was present in reactions without the helicase (10). There was much less synthesis in the reaction with 59 protein but no helicase (reaction 3, neutral gel, Fig. 6*B*), than in reaction 5 in which both the helicase and loading protein were omitted. The most likely explanation for this inhibition was that 59 protein bound tightly to the fork ahead of the leading strand polymerase, and inhibited polymerase progression (Fig. 6*A*). To test this hypothesis, we carried out the replication of the *ori(uvsY)* R loop plasmid in a two-stage reaction in which only the leading strand is labeled (Fig. 6*C*). The RNA was 3′ end-labeled by limited synthesis with dATP, dTTP, and P$^{32}$-dCTP in a first-stage reaction that required T4 polymerase, clamp, clamp loader, and 32 protein (10). 41 Helicase, 59 helicase loader, 61 primase, and topoisomerase then were added along with dGTP and enough unlabeled dCTP to prevent further $^{32}$P-dCTP incorporation. In the complete reaction most of the leading strand was full-length linear monomer on an alkaline gel (reaction 2, Fig. 6*C*). When 41

helicase, 59 helicase-loading protein, and primase were omitted, the remaining enzymes (polymerase, clamp, clamp loader, 32 protein, and topoisomerase) were sufficient for limited extension of the leading strand to about 2.5 kb (reaction 6). Helicase was loaded slowly without the loading protein (reaction 3). The length of the leading strand at early times was the same as that in reaction 6, in which helicase, helicase loader, and primase all were omitted. However, by 5 min there was some full-length linear monomer in the reaction with the helicase but no 59 protein (reaction 3). Dramatically, the leading strand elongation stopped at about 300–500 b, when 59 protein was present without the helicase (reaction 4). This result suggests that 59 protein, which as indicated above is a tight fork DNA binding protein (8, 9), binds ahead of polymerase at the replication fork and remains there until the helicase is loaded (Fig. 6*A*). From this position, the helicase-loading protein could help to coordinate leading and lagging strand synthesis by slowing the leading strand until helicase was available to work with primase to begin the lagging strand. It should be noted that in this experiment polymerase began elongating the RNA in the R loop in the first-stage reaction, and 59 protein was not added until the second stage. Thus, some or all of the 300–500 nt added when 59 protein was present without the helicase (reaction 4) may have been added by polymerase before 59 protein bound to the DNA.

The strong replication inhibition by 59 protein without the helicase with the supercoiled *ori(uvsY)* plasmid was much greater than its inhibition of rolling circle synthesis (16) (C. E. Jones and N.G.N., unpublished experiments). Possibly, the helicase-loading protein binds with higher affinity to a fork in a replication bubble on a supercoiled plasmid than to a fork on relaxed DNA.

The question of whether 59 protein blocks phage replication

*in vivo* in the absence of the helicase has not yet been addressed directly. However, it has been shown that wild-type T4 growth is inhibited in a host expressing 59 protein from a plasmid (36). This inhibition may be the result of replication blockage by 59 protein present at high levels at the time of infection.

**A DUE Is Required for Replication from an R Loop.** Because T4 41 helicase and 59 helicase-loading protein work on both replication forks and recombination intermediates, they must be able to bind at many sequences. Our recent experiments with R loop plasmids demonstrate that the requirement for loading the T4 replication enzymes, including the helicase and 59 helicase-loading protein, is not a specific T4 sequence, but rather a specific structure, an R loop within an easily unwound DNA sequence (DUE). The minimal T4 origins *ori(uvsY)* and *ori(34)* each contain AT-rich DUE sequences downstream of a T4 middle promoter. Deletion analysis showed that plasmid replication in a T4-infected cell gradually decreased as more of the DUE sequence was removed (31). A plasmid whose only T4 DNA was the *ori(uvsY)* middle promoter (pY074Δ) replicated poorly. However, replication could be restored by replacing the T4 DUE with a well characterized DUE from the pBR322 plasmid (32). Consistent with these results we found that although a stable R loop could be formed downstream of the T4 middle promoter on pY074Δ, this R loop was not extended by the T4 replication proteins *in vitro* (10), because of the absence of a DUE sequence. Furthermore, although pBR322 cannot be replicated in a T4-infected cell because it lacks the T4 promoter necessary to make RNA (32), pBR322 with a preformed R loop in the DUE region was replicated *in vitro* as well as the plasmid pKK405 with the R loop at T4 *ori(uvsY)* (10). These experiments explain earlier observations that origin function was retained when the T4 middle promoter at *ori(uvsY)* was replaced by other T4 middle promoter sequences, but middle promoters without an adjacent DUE are not origins (32).
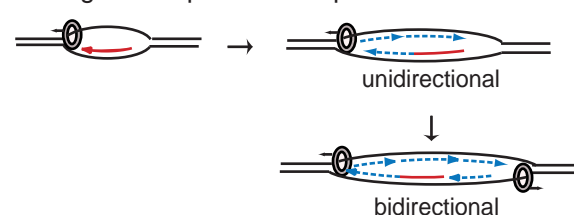
## The Versatility of T4 41 Helicase and 59 Helicase-Loading Protein Couples Replication and Recombination

T4 59 protein can bind and load the 41 helicase on forked DNA without sequence specificity as long as there are 10 single-stranded nt on the lagging strand template at the fork. It also can load the helicase on R loops if they are within easily unwound sequences, and on D loops, cruciforms, and strand invasion structures. This versatility allows T4 replication forks to be established on R loops created by transcription and at D loops created by recombination (Fig. 7).
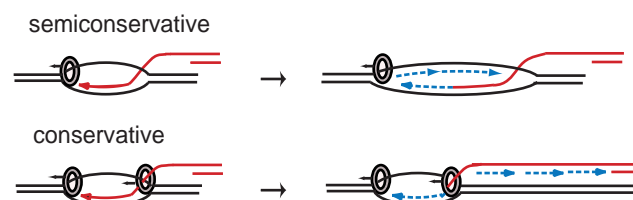
In a T4-infected cell, the origin R loops are formed by transcription using *E. coli* RNA polymerase modified for middle transcription in the case of *ori(uvsY)*, *ori(34)*, and *ori A* and unmodified host polymerase at *ori E* (3, 7, 31). Once the R loops are formed within DUEs, polymerase can be clamped at the 3′ end of the RNA by the gene 45 clamp and 44/62 clamp loader. 59 Protein can bind at the fork in front of the RNA and position the helicase on the 32 protein-covered displaced strand, without additional phage or host proteins (Fig. 7*A*). From this position helicase unwinds the duplex ahead of the leading strand polymerase and binds primase to begin lagging strand synthesis. Addition of topoisomerase is then sufficient for unidirectional replication from the origin. Bidirectional replication is established when helicase is loaded on the strand displaced by elongation of the first lagging strand fragment.

D loops and strand invasion structures are formed at double-stranded breaks and at the ends of the T4 linear genome by the T4-encoded uvsX recombinase, loaded with the help of the uvsY and 32 proteins (see refs. 2, 4, and 38 for recent reviews) (Fig. 7 *B* and *C*). Unidirectional semiconservative replication from a D loop requires a single helicase and is mechanistically similar to unidirectional replication from an origin R loop. At present there is no evidence indicating whether assembly of the T4 replication enzymes
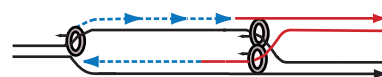


**Fig. 7.** Models for the role of 41 helicase in origin-initiated and recombination-initiated replication. Helicase hexamers (circles) are shown on the strands they surround. Arrows above the circles show the direction of helicase movement. Arrowheads indicate the 3′ strand ends. RNA and invading DNA molecules are in red. Newly synthesized DNA is a dashed line in blue. (*A*) One or two helicases are required, respectively, for unidirectional or bidirectional replication from an origin. The helicase at each fork opens the duplex ahead of the polymerase and enables primer synthesis by the primase. (*B*) Semiconservative and conservative replication at forks initiated by recombination. In semiconservative replication, the displaced strand serves as the lagging strand template, and the helicase role is like that at an origin. In conservative replication, the invading strand serves as the lagging strand template. The helicase at the right catalyzes the branch migration needed for reannealing the duplex behind the leading strand and also interacts with the primase on the invading strand. (*C*) Semiconservative replication coupled to branch migration at a four-way junction. Helicase at the fork opens the duplex and interacts with primase. Branch migration is catalyzed by 5′ to 3′ movement of helicase on the four-way junction behind the leading strand polymerase. Two helicase hexamers are shown on the junction because the products of unwinding a cruciform with nonhomologous arms are most easily explained by two 41 helicases unwinding the cruciform simultaneously (see Fig. 4). However, it remains to be determined whether a single hexamer is sufficient for strand exchange at a Holliday junction. See text for further discussion. Adapted from ref. 9.

at these D loops is enhanced by an easily unwound sequence, as has been shown for replication from R loops.

A second helicase would be required if conservative replication followed strand invasion (Fig. 7*B*). While the helicase at the left opened the duplex ahead of the leading strand polymerase elongating the invading strand, the helicase at the right would move on the invading strand in the same direction as the polymerase, unwinding the newly synthesized strand. Interaction between the second 41 helicase and 61 primase would allow synthesis of the primers needed to initiate discontinuous synthesis on the unwound portion of the newly synthesized strand. In the absence of primase, helicase traveling behind polymerase on the newly synthesized strand leads to bubble branch migration, as first observed in the T4 system with the Dda helicase by Formosa and Alberts (37). Although 59 protein can bind and load helicase on an invading strand (Fig. 4 and ref. 9), at least two factors favor semiconservative rather than conservative replication after strand invasion. First, the helicase on the invading strand would be less stable than the helicase ahead of the

COLLOQUIUM

polymerase, because there is evidence suggesting that interaction between 41 helicase and T4 DNA polymerase helps to keep the helicase on the fork (36, 39). Second, uvsX protein coating the invading strand would inhibit binding by the 59 helicase-loading protein (13, 38). Conservative replication with the T4 proteins has not been demonstrated *in vivo* or *in vitro*.

T4 59 helicase-loading protein, 41 helicase, and 32 protein together catalyze strand exchange (27, 28) and unwind four-stranded Holliday junctions (9). Recombination-initiated semiconservative replication can be coupled to branch migration if 41 helicase binds to the four-way junction behind the leading strand polymerase (Fig. 7C) and moves the junction toward the replication fork. The T4 helicase responsible for semiconservative replication must be 41 helicase, because it is the only helicase that interacts with the primase. T4 Dda helicase could replace 41 helicase at the four-way junction. These coupled replication and recombination reactions would result in coordinated extension of the two strands of the invading DNA, leaving the donor DNA unchanged, as recently proposed for the break-induced replication model for double-strand break repair (40, 41). This replication mechanism with helicase at the fork coordinating leading and lagging strand synthesis, and helicase behind the fork catalyzing strand exchange, is well suited to the T4-infected cell where most replication begins with recombination. It also may account for the extensive chromosomal replication (termed ECR) that has been shown to accompany double-strand break repair in T4 phage DNA (2, 42).

## Comparison with Other Helicases and Loading Proteins

T4 gene 59 helicase-loading protein differs from other proteins with this function in being able to load a helicase on origins, on replication forks remote from origins, and on three- and four-stranded replication intermediates. Separate sets of proteins are required for these multiple functions in *E. coli*. Like T4 gene 41 helicase, *E. coli* dnaB helicase can slowly load on forked DNA

by itself (43). However, both the dnaA protein, which binds to dnaA box sequences on *oriC*, and dnaC protein, which forms a hexameric complex with the dnaB hexamer, are required to load the dnaB helicase at this origin. The O and P proteins of phage lambda are required to load the dnaB helicase at the phage origin (reviewed in ref. 44). In contrast to 59 protein, the single-stranded DNA binding activities of the lambda P and *E. coli* dnaC proteins are only evident when these proteins form complexes with the dnaB helicase (45).

*E. coli* Pri A protein shares with T4 59 protein the ability to bind with high affinity to forked DNA and to the three-stranded junction at the 3′ end of an invading strand in D loops, but Pri A does not bind to four-stranded Holliday junctions. PriA is needed to load the dnaB helicase on stalled replication forks (46–49), on D loops at specialized origins during inducible stable replication, on R loops during constitutive stable replication in *E. coli* RNase HI mutants (50, 51), and on the phage Mu strand transfer complex (52, 53). Thus PriA resembles T4 59 in recognizing forks and D loops without sequence specificity, but it does not bind or load dnaB helicase on four-stranded Holliday junctions.

In *E. coli* a separate helicase (RuvB) and helicase-loading protein (RuvA) are responsible for branch migration on recombination structures. The RuvAB complex operates by a mechanism distinct from that of the replicative helicases, and it has no role in leading or lagging strand synthesis (54–56). In the simpler bacteriophage T4 system, 41 helicase, loaded by 59 protein, can catalyze branch migration and is essential for both origin- and recombination-initiated replication.

1. Kreuzer, K. N. & Morrical, S. W. (1994) in *Molecular Biology of Bacteriophage T4*, ed. Karem, J. (Am. Soc. Microbiol., Washington, DC), pp. 28–42.
2. Kreuzer, K. N. (2000) *Trends Biochem. Sci.* **25,** 165–173.
3. Mosig, G., Colowick, N., Gruidl, M. E., Chang, A. & Harvey, A. J. (1995) *FEMS Microbiol. Rev.* **17,** 83–98.
4. Mosig, G. (1998) *Annu. Rev. Genet.* **32,** 379–413.
5. Nossal, N. G. (1994) in *Molecular Biology of Bacteriophage T4*, ed. Karem, J. (Am. Soc. Microbiol., Washington, DC), pp. 43–53.
6. Baker, T. A. & Bell, S. P. (1998) *Cell* **92,** 295–305.
7. Vaiskunaite, R., Miller, A., Davenport, L. & Mosig, G. (1999) *J. Bacteriol.* **181,** 7115–7125.
8. Mueser, T. C., Jones, C. E., Nossal, N. G. & Hyde, C. C. (2000) *J. Mol. Biol.* **296,** 597–612.
9. Jones, C. E., Mueser, T. C. & Nossal, N. G. (2000) *J. Biol. Chem.* **275,** 27145–27154.
10. Nossal, N. G., Dudas, K. C. & Kreuzer, K. N. (2001) *Mol. Cell* **7,** 31–41.
11. Wu, R., Wu, J. L. & Yeh, Y. C. (1975) *J. Virol.* **16,** 5–16.
12. Shah, D. B. (1976) *J. Virol.* **17,** 175–182.
13. Yonesaki, T. (1994) *J. Biol. Chem.* **269,** 1284–1289.
14. Morrical, S. W., Hempstead, K. & Morrical, M. D. (1994) *J. Biol. Chem.* **269,** 33069–33081.
15. Morrical, S. W., Beernink, H. T. H., Dash, A. & Hempstead, K. (1996) *J. Biol. Chem.* **271,** 20198–20207.
16. Barry, J. & Alberts, B. (1994) *J. Biol. Chem.* **269,** 33049–33062.
17. Zlatanova, J. & van Holde, K. (1998) *FASEB J.* **12,** 421–431.
18. Pöhler, J. R., Norman, D. G., Bramham, J., Bianchi, M. E. & Lilley, D. M. (1998) *EMBO J.* **17,** 817–826.
19. Hill, D. A., Pedulla, M. L. & Reeves, R. (1999) *Nucleic Acids Res.* **27,** 2135–2144.
20. Hardman, C. H., Broadhurst, R. W., Raine, A. R., Grasser, K. D., Thomas, J. O. & Laue, E. D. (1995) *Biochemistry* **34,** 16596–16607.
21. Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R. & Wright, P. E. (1995) *Nature (London)* **376,** 791–795.
22. Fenley, M. O., Manning, G. S., Marky, N. L. & Olson, W. K. (1998) *Biophys. Chem.* **74,** 135–152.
23. Raney, K. D., Carver, T. E. & Benkovic, S. J. (1996) *J. Biol. Chem.* **271,** 14074–14081.
24. Cunningham, R. P. & Berger, H. (1978) *Virology* **88,** 62–70.
25. Yonesaki, T. (1994) *Genetics* **138,** 247–252.
26. Kuroki, E. & Yonesaki, T. (1999) *Mol. Gen. Genet.* **262,** 525–533.
27. Salinas, F. & Kodadek, T. (1995) *Cell* **82,** 111–119.
28. Kong, D., Nossal, N. G. & Richardson, C. C. (1997) *J. Biol. Chem.* **272,** 8380–8387.
29. Mosig, G. & Colowick, N. (1995) *Methods Enzymol.* **262,** 587–604.
30. Kreuzer, K. N., Yap, W. Y., Menkens, A. E. & Engman, H. W. (1988) *J. Biol. Chem.* **263,** 11366–11373.
31. Menkens, A. E. & Kreuzer, K. N. (1988) *J. Biol. Chem.* **263,** 11358–11365.
32. Carles-Kinch, K. & Kreuzer, K. N. (1997) *J. Mol. Biol.* **266,** 915–926.
33. Stitt, B. & Hinton, D. (1994) in *Molecular Biology of Bacteriophage T4*, ed. Karam, J. D. (Am. Soc. Microbiol., Washington, DC), pp. 142–160.
34. Belanger, K. G. & Kreuzer, K. N. (1998) *Mol. Cell* **2,** 693–701.
35. Minden, J. S. & Marians, K. J. (1986) *J. Biol. Chem.* **261,** 11906–11917.
36. Spacciapoli, P. & Nossal, N. G. (1994) *J. Biol. Chem.* **269,** 447–455.
37. Formosa, T. & Alberts, B. M. (1986) *Cell* **47,** 793–806.
38. Beernink, H. T. & Morrical, S. W. (1999) *Trends Biochem. Sci.* **24,** 385–389.
39. Dong, F., Weitzel, S. E. & von Hippel, P. H. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 14456–14461.
40. Holmes, A. M. & Haber, J. E. (1999) *Cell* **96,** 415–424.
41. Haber, J. E. (1999) *Trends Biochem. Sci.* **24,** 271–275.
42. George, J. W. & Kreuzer, K. N. (1996) *Genetics* **143,** 1507–1520.
43. Stephens, K. M. & McMacken, R. (1997) *J. Biol. Chem.* **272,** 28800–28813.
44. Baker, T. A. & Wickner, S. H. (1992) *Annu. Rev. Genet.* **26,** 447–477.
45. Learn, B. A., Um, S. J., Huang, L. & McMacken, R. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 1154–1159.
46. Liu, J. & Marians, K. J. (1999) *J. Biol. Chem.* **274,** 25033–25041.
47. Sandler, S. J. & Marians, K. J. (2000) *J. Bacteriol.* **182,** 9–13.
48. Marians, K. J. (2000) *Trends Biochem. Sci.* **25,** 185–189.
49. Marians, K. J. (2000) *Curr. Opin. Genet. Dev.* **10,** 151–156.
50. Masai, H., Asai, T., Kubota, Y., Arai, K. & Kogoma, T. (1994) *EMBO J.* **13,** 5338–5345.
51. Kogoma, T. (1997) *Microbiol. Mol. Biol. Rev.* **61,** 212–238.
52. Jones, J. M. & Nakai, H. (1999) *J. Mol. Biol.* **289,** 503–516.
53. Jones, J. M. & Nakai, H. (2000) *Mol. Microbiol.* **36,** 519–527.
54. Parsons, C. A., Tsaneva, I., Lloyd, R. G. & West, S. C. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 5452–5456.
55. West, S. C. (1997) *Annu. Rev. Genet.* **31,** 213–244.
56. Yu, X., West, S. C. & Egelman, E. H. (1997) *J. Mol. Biol.* **266,** 217–222.
57. Kreuzer, K. N. & Alberts, B. M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 3345–3349.